CALS DATABASE USAGE AND ANALYSIS TOOL STUDY
FINAL REPORT

September 1991

Contract MDA903-90-D-0022

Prepared for:

Defense Logistics Agency
Room 3C529
Cameron Station
Alexandria, VA 22304-6100

Prepared by:

James H. Cook
Edward J. Szwedo

IIT Research Institute
201 Mill Street
Rome, NY 13440-2069

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave Blank*)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1991	Final Report Oct 1990 - Sept 1991

4. TITLE AND SUBTITLE	5. FUNDING NUMBERS
CALS Database Usage Analysis Tool Study	MDA903-90-D-0022

6. AUTHOR(S)

James H. Cook
Edward J. Szwedo

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	8. PERFORMING ORGANIZATION REPORT NUMBER
IIT Research Institute 201 Mill Street Rome, NY 13440	

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING AGENCY REPORT NUMBER
Defense Logistics Agency (DLA/ZIR) Room 3C529 Cameron Station Alexandria, VA 22304-6100	

11. SUPPLEMENTARY NOTES:

12a. DISTRIBUTION/AVAILABILITY STATEMENT	12b. DISTRIBUTION CODE
Approved for public release; distribution unlimited	Unclassified

13. ABSTRACT (Maximum 200 words)

The computer Assisted Acquisition and Logistics Support (CALS) effort will involve large and geographically dispersed databases of proprietary technical information pertaining to weapons systems and parts. These databases will be combined to form the CALS Integrated Weapons Systems Database (IWSDB). While a goal of the CALS effort is to openly exchange information in a standardized format, it is recognized that unlimited access to large amounts of weapons systems data will pose a security risk. These databases will be shared by government and industry personnel and by their very nature will contain data that when aggregated could increase in sensitivity or classification. The purpose of this study was to develop a conceptual design for a tool that will monitor CALS database usage in real-time to prevent unauthorized access to potentially sensitive and proprietary data. 25 * Com[illegible] - [illegible] instruction

14. SUBJECT TERMS	15. NUMBER OF PAGES
Data Aggregation, Database Security. * Logistics Support, * Data bases, * Data acquisition	19
	16. PRICE CODE

17. SECURITY CLASSIFICATION	18. SECURITY CLASSIFICATION	19. SECURITY CLASSIFICATION	20. LIMITATION OF ABSTRACT
UNCLASSIFIED			

CALS DATABASE USAGE AND ANALYSIS TOOL STUDY
FINAL REPORT

September 1991

Contract MDA903-D-0022

Prepared for:

Defense Logistics Agency
Room 3C529
Cameron Station
Alexandria, VA 22304-6100

Prepared by:

James H. Cook
Edward J. Szwedo

IIT Research Institute
201 Mill Street
Rome, NY 13440-2069

EXECUTIVE SUMMARY

The Computer Assisted Acquisition and Logistics Support (CALS) effort will involve large and geographically dispersed databases of proprietary technical information pertaining to weapons systems and parts. These databases will be combined to form the CALS Integrated Weapons Systems Database (IWSDB). While a goal of the CALS effort is to openly exchange information in a standardized format, it is recognized that unlimited access to large amounts of weapons systems data will pose a security risk. These databases will be shared by government and industry personnel and by their very nature, will contain data that, when aggregated, could increase in sensitivity or classification.

The data aggregation problem, as it applies to non-distributed databases, is a current topic of ongoing research in computer database security. The purpose of this study was to develop a conceptual design for a tool that will monitor CALS database usage in real-time to prevent unauthorized access to potentially sensitive and proprietary data.

IITRI began this effort by investigating ongoing research on data aggregation. The results of this literature search were documented in a bibliography that contains citations for 38 documents. The bibliography is included as Appendix A.

During the second phase of this effort, IITRI developed a Requirements Analysis Report. Our requirements analysis indicated that a Data Aggregation Tool (DAT) should take into consideration the *prevention* of aggregation as well as *detection* and recovery from compromise. Towards prevention, a DAT should be capable of protecting sensitive relationships between data items. Towards detection, the tool should operate in polynomial time and support a variable audit capability. With regards to recovery, a DAT should include a learning element that will derive new security rules to prevent a compromising scenario from recurring. To this end, the DAT should be designed in conjunction with the IWSDB to ensure a synergistic relationship. An object-oriented DBMS was found to hold much promise for handling the aggregation problem. The Requirements Analysis Report is included as Appendix B.

In this, the Final Report, the project team developed a model of the aggregation process to identify aspects of the process which might be exploited by the DAT. To illustrate the process of aggregation an analogy to the process of assembling a jigsaw puzzle was developed. Using the analogy as a guide, a model of the aggregation process was developed. The results of the modeling process were then used to develop various concepts which should be included in the design of a Data Aggregation Tool. These concepts include a database analysis and design tool, a database usage monitoring tool, and a learning tool.

A CONCEPT FOR ADDRESSING AGGREGATION

1.0 INTRODUCTION

The Computer Assisted Acquisition and Logistics Support (CALS) effort will involve large and geographically dispersed databases of proprietary technical information pertaining to weapons systems and parts. These databases will be combined to form the CALS Integrated Weapons Systems Database (IWSDB). While a goal of the CALS effort is to openly exchange information in a standardized format, it is recognized that unlimited access to large amounts of weapons systems data will pose a security risk. These databases will be shared by government and industry personnel, and by their very nature will contain data that, when aggregated, could increase in sensitivity or classification.

The objective of this effort is to develop a conceptual design for a tool that will help to prevent the aggregation of sensitive or classified information from unclassified components of that information. A previous report discussed high level requirements for a Data Aggregation Tool (DAT). This final report presents the major ideas developed for the conceptual design of a DAT.

Early in the effort the project team decided that a model of the aggregation process would be useful to identify aspects of the process which might be exploited by the DAT. To illustrate the process of aggregation an analogy to the process of assembling a jigsaw puzzle was developed.

This report first introduces the process of aggregation by describing this analogy. Then, using the analogy as a guide, a model of the aggregation process is developed. The concluding sections use the results of the model and the requirements identified earlier to develop various concepts which should be included in the design of a Data Aggregation Tool. For instance we will show that careful database design principles can be employed to significantly reduce the possibility of aggregation.

2.0 THE JIGSAW PUZZLE ANALOGY FOR AGGREGATION

One problem that arises in discussing the process of aggregation is the lack of a concrete example of aggregation. A straightforward description might be "assembling enough unclassified *things* in a database to enable some classified information to be discovered" but it is too abstract and the reader is unlikely to develop an intuitive feeling for the process. The jigsaw puzzle analogy described below overcomes this problem by explaining aggregation in terms of a common activity most people have participated in.

The analogy proceeds as follows: Suppose one is given a number of disassembled jigsaw puzzles all mixed together in a bin. The bin represents a multilevel secure database, and the individual pieces represent records in the database. The image fragments on each piece represent the information in that record. The image fragments themselves are unclassified, but the information represented by the picture on one of the assembled puzzles is classified. The process of picking enough of the pieces belonging to the puzzle having a classified picture, assembling these pieces, and thereby discovering the classified subject represents a security compromise by aggregation.

It may not, however, be necessary to assemble the entire puzzle to determine the picture. There may be one or more key subsets of puzzle pieces, such as the smile on the Mona Lisa, that, when recognized, allow one to infer something about the picture as a whole, (e.g. the picture is of a person, that person may be the Mona Lisa). The notion of key subsets will be discussed further below.

3.0 MODELING THE AGGREGATION PROCESS

Compromising the security of data in a multilevel database by the use of aggregation can be thought of as a process. Our approach to developing a concept for defeating aggregation is to model this process and quantify the impact of various parameters on the probability of the process succeeding. The conceptual design of a DAT will then exploit those parameters having the greatest impact.

3.1 Introducing the Model

In this section we develop a model of the aggregation process following the jigsaw puzzle analogy presented above. Figure 1 is a flow diagram of the model. The various activities represented in the model do not necessarily occur in chronological order. They represent activities which may be performed simultaneously.

The subject data is represented by a bin containing pieces from N jigsaw puzzles. One of these N puzzles is the target. The remainder can be considered irrelevant data. The puzzle pieces are all uniformly shaped and colored on the back so that one cannot easily tell those which belong to the same puzzle. The pattern on the front of the puzzle is classified in the aggregate. Individual pieces are unclassified. The pattern on the front of each piece may provide a subtle cue linking pieces to a particular puzzle. In addition, the pattern on the front of the puzzles will provide unique matches for any puzzle pieces which should be connected.[1]

[1] In Figure 1, for purposes of illustration, the target pieces are represented by black squares and the irrelevant data by striped squares.

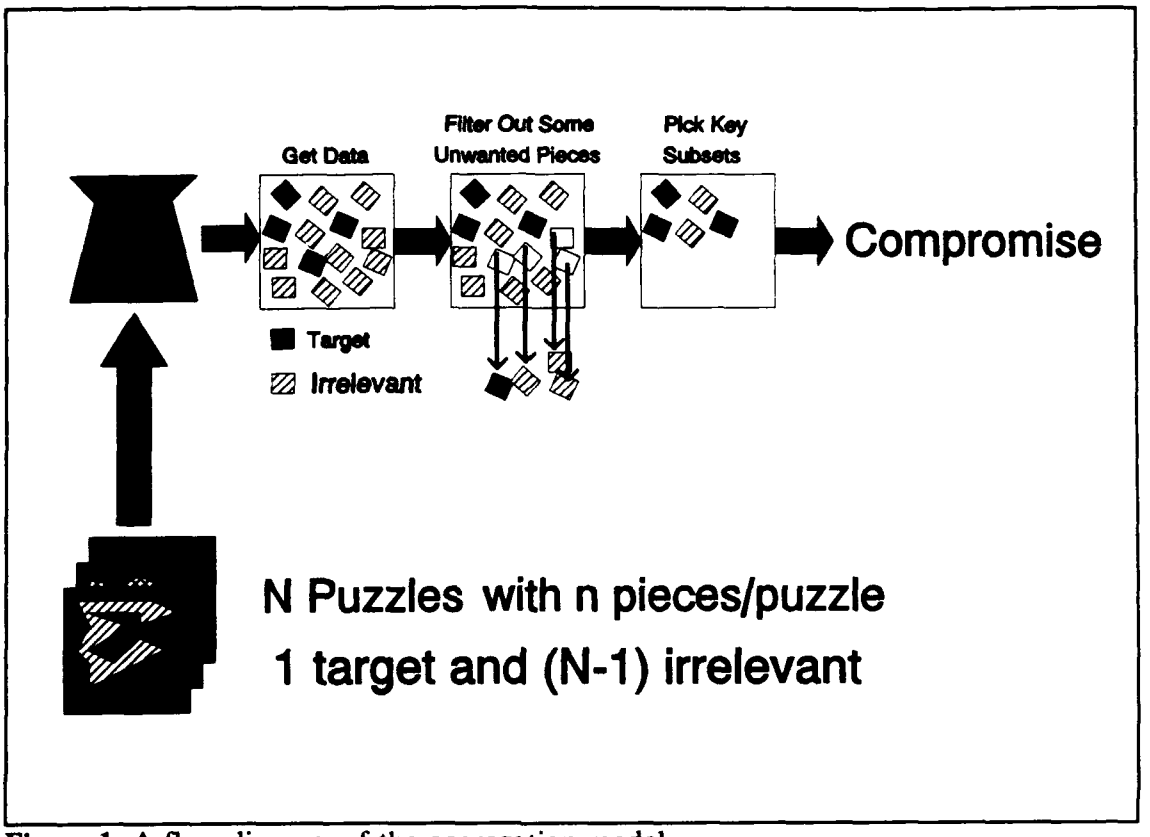

Figure 1. A flow diagram of the aggregation model.

The subject data, i.e. bin of pieces, acts as input to the first activity which involves getting the data. This would be the process of selecting a subset of puzzle pieces and laying them out for examination.

The next activity involves attempting to select the pieces of a single target puzzle by filtering out undesired pieces. This may be easy if the backs of all pieces from a given puzzle have an identifying marking, color, or shape. If they don't, as is the case here, it would be quite difficult.

After the selection and filtering activities, the pieces would be passed on to an activity in which potential key subsets of puzzles would be assembled. A group of pieces would be locked together (we assume that all puzzles have similarly shaped pieces that can be interchanged) and correlated with outside data. Essentially this involves looking at the picture fragment and asking oneself if that fits with anything in one's memory. For instance, if the fragment is recognized as part of an eye one might surmise that the picture is of a face. If it is a three digit number enclosed in parenthesis followed by another three digit number, one might consider that it could be part of a telephone number.

If the correlation is successful and the entire picture is guessed the process ends in success. If the correlation isn't successful, i.e. the subset identified isn't a key subset, then the process is defeated.

We assume that this is a one-pass process. One could return to the bin after each failure and eventually be sure of getting a key subset. An equivalent process would be to initially take enough pieces from the bin to get this key subset. Since a multiple pass process complicates our model, we chose to consider only the one-pass process.

Each of the activities of this process has a particular impact on the probability of the entire process succeeding. In the following sections we discuss and quantify the impact of these activities. Our goal is to establish an expression for the probability of the process succeeding. In addition to identifying the number of pieces necessary to compromise security, the expressions we develop will enable us to identify the most useful parameters to exploit in the development of a DAT.

3.2 The Subject Data

To properly represent the real world, the subject database must contain data at two levels of classification. For each of the N puzzles, most of the pieces will be at level 0, i.e. unclassified, but a certain fraction of pieces from each puzzle may be classified at level 1. The information represented by the picture or image printed on each puzzle will be classified at level 1.

We assume that user access will be restricted to level 0, since we are modeling the process of compromising the level 1 information.

3.3 The Data Filtering Activity

Data filtering impacts the aggregation process by increasing the probability of obtaining key subsets, thus enabling one to concentrate pieces from the target puzzle. This impact can be quantified by determining the probability, P_L, that the filter will provide at least an L% pure data set, i.e., provide mL target pieces. This probability will depend upon

- L: the purity of the data set - (expressed as a percentage such that mL is an integer)
- m: the size of the data set selected,
- N: the number of different puzzles,
- n: the number of pieces in a typical puzzle, and
- c: the quality of the available cues which help indentify those pieces belonging to common puzzles

Let c be a number between 0 and N such that the probability of picking a target puzzle piece from the bin of puzzle pieces is c/N. If there are no cues the probability of picking the target piece is determined purely by chance, and c would be equal to 1 (assuming n>>1). If there are negative cues, i.e. deception, the chance of picking a target piece would be less than that of pure chance and c would be less than 1. Positive cues would have c between 1 and N.[2]

The probability P_L is then easily shown to be given by the following formula assuming n is large, which would probably be the case for the IWSDB (i.e. the number of unclassified parts comprising a potentially sensitive subsystem would be large).

$$P_L = (c/N)^{mL} (1 - c/N)^{(m-mL)} \frac{m!}{(mL)!(m-mL)!} \quad (1)$$

This may be transformed into the binomial probability by substitution as follows:

$$\lambda = \frac{cm}{N} \quad (2)$$

$$P_L = (\lambda/m)^{mL} (1 - \lambda/m)^{m-mL} \frac{m!}{(mL)!\,(m-mL)!} \quad (3)$$

If m > 20 and N > 20c, which is not unreasonable for the CALS IWSDB, the probability of obtaining an L% pure data set may be approximated by the Poisson distribution,

$$P_L \sim \frac{\lambda^{mL} e^{-\lambda}}{(mL)!} \quad (4)$$

---

[2] Note that there will be other information associated with each piece which links it to other pieces of the same puzzle. This information will be classified at level 1 to enable cleared users to get the information printed on the puzzle. The cues in this model are different from this information. They only statistically improve the chances of linking a piece to the target puzzle, and they are not classified. Compromising the classified information by gaining access to the level 1 information linking target puzzle pieces is not an aggregation problem.

$$\textit{with mean and variance} = \lambda = \frac{cm}{N} \tag{5}$$

Thus, if c is small and N and m are large enough the mean and variance of mL equal cm/N, i.e.

$$\langle L \rangle = \frac{c}{N} \tag{6}$$

$$\langle (L - \langle L \rangle)^2 \rangle = \frac{c}{mN} \tag{7}$$

This seems intuitive. If there were 20 puzzles and there were no cues, one would expect that 1 out of every 20 draws would be from the target puzzle on the average. In addition, as expected, the standard deviation of L is inversely proportional to the square root of the number of samples obtained.

A few examples of P_L for various parameters are provided in Figure 2. We can see from Figure 2 that the mean of L increases with c.

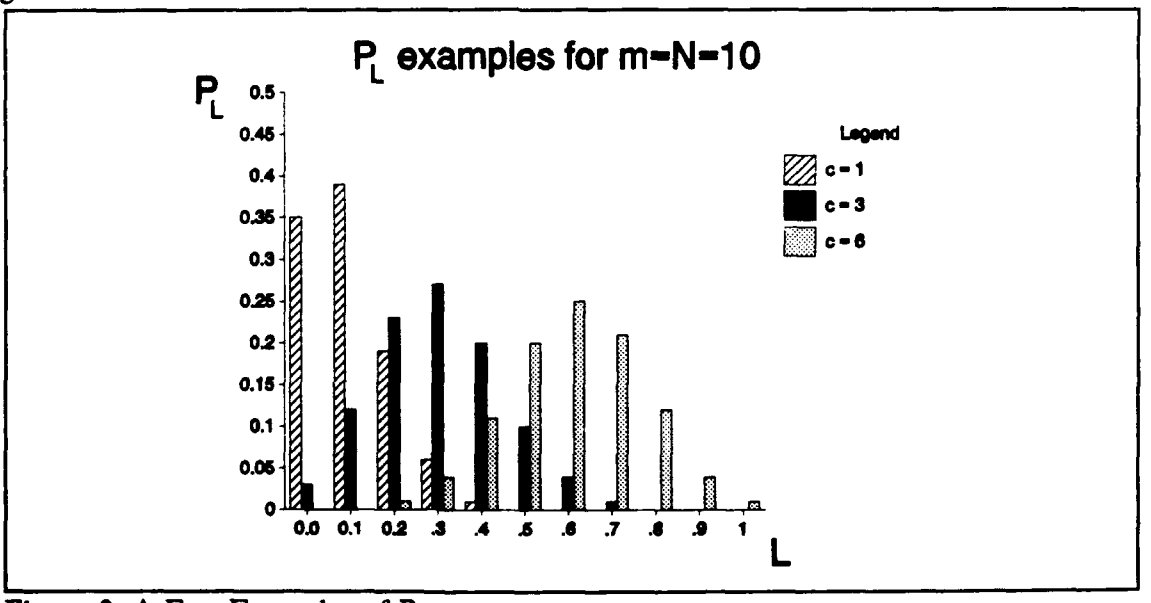

Figure 2. A Few Examples of P_L.

3.3 Picking Key Subsets

We assume that once two adjacent puzzle pieces are obtained, the join will be made with a certainty of 1 using the information side of the puzzle pieces to show the match. Here we are trying to quantify the number of pieces that are needed to develop enough joins in enough key subsets to compromise the information side of the puzzle.

Let K_1 = { the smallest key subset of pieces }
K_2 = { the next larger key subset of pieces }
K_3 = { the next larger key subset of pieces }
etc. for K_4, ...K_i ...

Further, let α_i be the fraction of K_i needed to be able to identify key subset K_i and compromise the information on the front of the puzzle.

The approach to soiving the problem is to develop the probability, $P_{\alpha imL}$ that mL pieces from the target puzzle will provide $\alpha_i K_i$ pieces for the i^{th} key subset. Combining the results for all of the most significant key subsets provides P_{mL}, the probability that mL pieces randomly picked will provide a necessary fraction of pieces from some key subset. Picking a conservative threshold for P_{mL} and combining it with a conservative estimate of L will provide a threshold for m, the number of pieces an uncleared user can pick without being likely to be able to compromise the information on the front of the puzzle.

$P_{\alpha imL}$ can be approximated for values of m and K_i that are small with respect to n as follows:

$$P_{\alpha imL} = \frac{K_i}{n} X \frac{K_i - 1}{n} X \ldots X \frac{K_i - (\alpha_i K_i - 1)}{n} X \frac{n - \alpha_i K_i}{n} X \frac{n - (\alpha_i K_i + 1)}{n} X \ldots X \frac{n - (mL - 1)}{n} X \frac{(mL)!}{(\alpha_i K_i)!\,(mL - \alpha_i K_i)!} \quad (8)$$

Figure 3 shows the general relative relationships for $P_{\alpha imL}$ for key subsets $K_1..K_3$ and the corresponding P_{mL}.

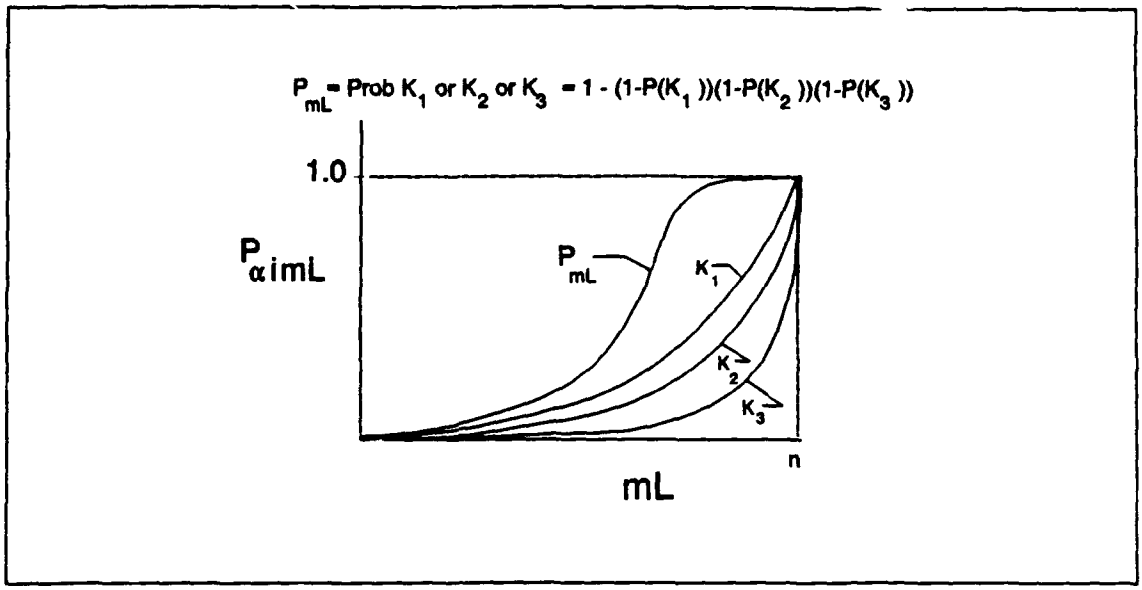

Figure 3. General Relationships for $P_{\alpha imL}$

4.0 DEVELOPING THE CONCEPT

At this point, we have developed qualitative expressions representing the impact of the activities in the model on the overall aggregation process. The next step is to develop an overall concept for a DAT which will exploit the insight provided by the model. In developing this concept we address the aggregation problem from three directions. First, we address issues dealing with prevention. Then, for those cases in which prevention may not be enough, we discuss detection mechanisms. Finally, for those cases in which compromise occurs, we discuss issues dealing with recovery.

4.1 Preventing Aggregation

The first line of defense in aggregation control is to implement a prevention mechanism. Using the jigsaw puzzle analogy this mechanism should prevent a user from assembling any key subsets from which classified information can be inferred.

Considering the model, this would involve taking measures to ensure that the purity, L, of a sample, m, and the probability of assembling a key subset given mL pieces from the target puzzle, P_{mL}, are kept as small as possible. As Figure 3 indicates, P_{mL} increases with mL and with the number of key subsets which can be compromised from a small numbers of pieces. Thus to prevent aggregation one could seek to

1. minimize m, the size of the data set selected,
2. minimize L, the purity of the data set, and
3. minimize the number of key subsets available to uncleared users.

Minimizing the Size of the Data Set

Minimizing m is an obvious measure which has been proposed in the literature. One approach is to require that any user needing a subset of data which may potentially lead to a security compromise via aggregation should have to obtain that data via a cleared user. While this would put an intelligent interface between the uncleared user and the data, it would make the entire data collection process so cumbersome that the price of protecting a key subset would be prohibitive.

In the context of a CALS Database this approach would not seem feasible. Thousands of people will need to access small amounts of data and much legitimate progress would be impeded by encumbering the process. In effect it would negate much of the advantage of computerizing the data in the first place.

Minimizing the Purity of the Data Set

Minimizing L would seem to show more promise. Equation 6 shows that <L> is proportional to the cue factor c and inversely proportional to the number of puzzles N. Both of these parameters can be controlled by careful database design. Essentially the goal would be to keep the user from assembling a large number of pieces from the target puzzle either by (1) recognizing some characteristic common to the pieces of the target puzzle or (2) relying on opportunities provided by chance. In the model this involves making c as small as possible (perhaps even less than one) and making N as large as possible.

The cue parameter c can be made small by making the membership of the data to common classes transparent to the user. This is analogous to making all the puzzle pieces, even from different puzzles, the same shape and color (as seen from the back). N can be made large by not isolating records according to common classes. For instance, do not keep all missile data physically located at nodes associated with missile contractors.

An object-oriented data model holds much promise for minimizing the cue parameter. The object model supports a hierarchy of classes of objects. Also built into the model is the capability to define relationships between any two levels of the hierarchy. Figure 4 illustrates objects at two different classification levels. Let object A represent a classified subsystem that is comprised of unclassified objects B through E. Rather than encumbering the low-level objects by classifying them, it is possible to classify only the "Is-Part-Of" relationship between the low-level and the high-level object.

As an illustration, consider the case where a collection of unclassified parts comprises a missile guidance system. We can protect the guidance system from being compromised through aggregation by classifying the relationship of the parts to the whole. Individuals accessing the missile guidance system parts who lack the clearance to access the guidance system itself would not even know of the relationship of the parts to the parent class, that

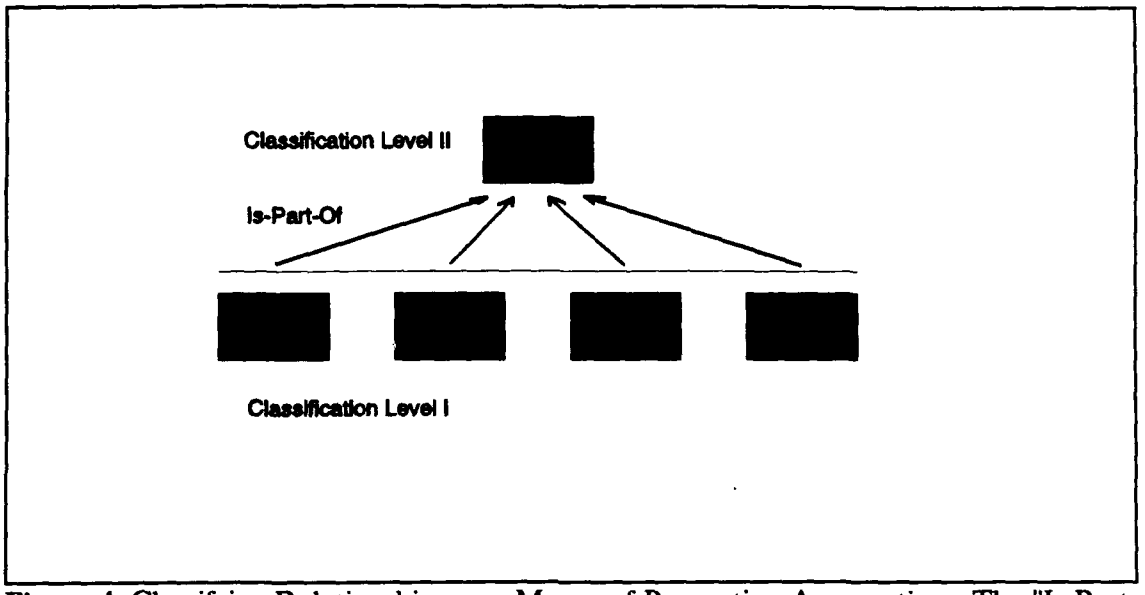

Figure 4. Classifying Relationships as a Means of Preventing Aggregation. The "Is-Part-Of" relationship and the data describing A are at Classification Level II. The data describing B, C, D, and E are at Classification Level I.

is, the guidance system. In this way many parts of a classified system can be made available to users (because the parts themselves are unclassified), and only those cleared to the appropriate level will be aware of the fact that the parts belong to the classified system.

Minimizing the Number of Key Subsets

Minimizing the number of small key subsets available to uncleared users is a challenge, particularly that aspect of the process involving the identification of such subsets. This process must be closely tied to the decisions made involving detection which will be discussed below. One way of minimizing the number of small key subsets is to classify them. Of course this inhibits access to unclassified information, but by their nature, small key subsets may be justifiably classed sensitive and should be treated accordingly.

The process of identifying key subsets should probably be carried out manually. One approach would be to assemble a small panel of cleared experts who are all familiar with the CALS IWSDB. This group would attempt to identify all potential small key subsets. Then a larger panel of cleared users, independent of the first panel would be assembled. A Monte Carlo experiment would be performed on this group by showing them each random parts of each potential key subset and having them guess the target information. The results of such experiments can provide an empirical probability distribution such as shown in Figure 3. As we gain experience through repeated experiments it may be possible to identify various classes of information represented in the key subsets which seem to

behave similarly with respect to the probability of compromise. Once such classifications are known the process can probably be partially automated.

Summarizing, the concepts most important to the prevention of aggregation appear to involve the careful design of the database to keep the availability of concentrated subsets from the target data low and the classification of all small key subsets belonging to the target data. Taken together they would minimize the probability of aggregation for small subsets of data. Below, we discuss establishing a threshold for m to help quantify this minimum probability and the size of a *small* subset.

4.2 Detecting Aggregation

A DAT must also help to detect aggregation in those cases where prevention fails. One approach to detection is to track, for each user, the number of records obtained, sorted by key subset. When the user approaches a subset's limit regarding the amount of data retrieved, a parameter would be set to prevent any further accesses by the user and the transaction would be flagged for investigation. Such a process is straightforward. The limits associated with each subset could be developed during the process discussed above to identify small key subsets. However, given the size and complexity of the IWSDB, the large number of users, and the realtime requirements for detection, this process would be a considerable burden on the system.

A much simpler variation would involve only putting a limit on the total number of records each user accesses, the parameter m in the jigsaw model. By extending the process mentioned above for identifying the key subsets, the distribution, P_{mL}, can be established. As illustrated in Figure 3, P_{mL} represents the probability for compromising any of the key subsets identified.

Once P_{mL} is established, using the probability for L (P_L) given in equation 1 and estimating a discrete probability distribution for m (P_m), the overall probability of compromise can be written.

$$P_{comp} = \sum_{m=1}^{m_t} \sum_{L=0,1/m,..}^{1} P_{mL} P_L P_m \qquad (9)$$

As noted, this assumes that a limit, m_t, is set on the number of records a user can obtain. Figure 5 contains a plot of P_{comp} as a function of m_t assuming a uniform density for P_m between 1 and m_t, P_L as given in equation 1, and P_{mL} represented as a quadratic in mL such that $P_{mL} = 0.2$ when $mL = 30$.

Thus the concept for detecting aggregation would involve, for a class of N data sets, estimating the various probabilities in equation 9 and determining the overall probability of

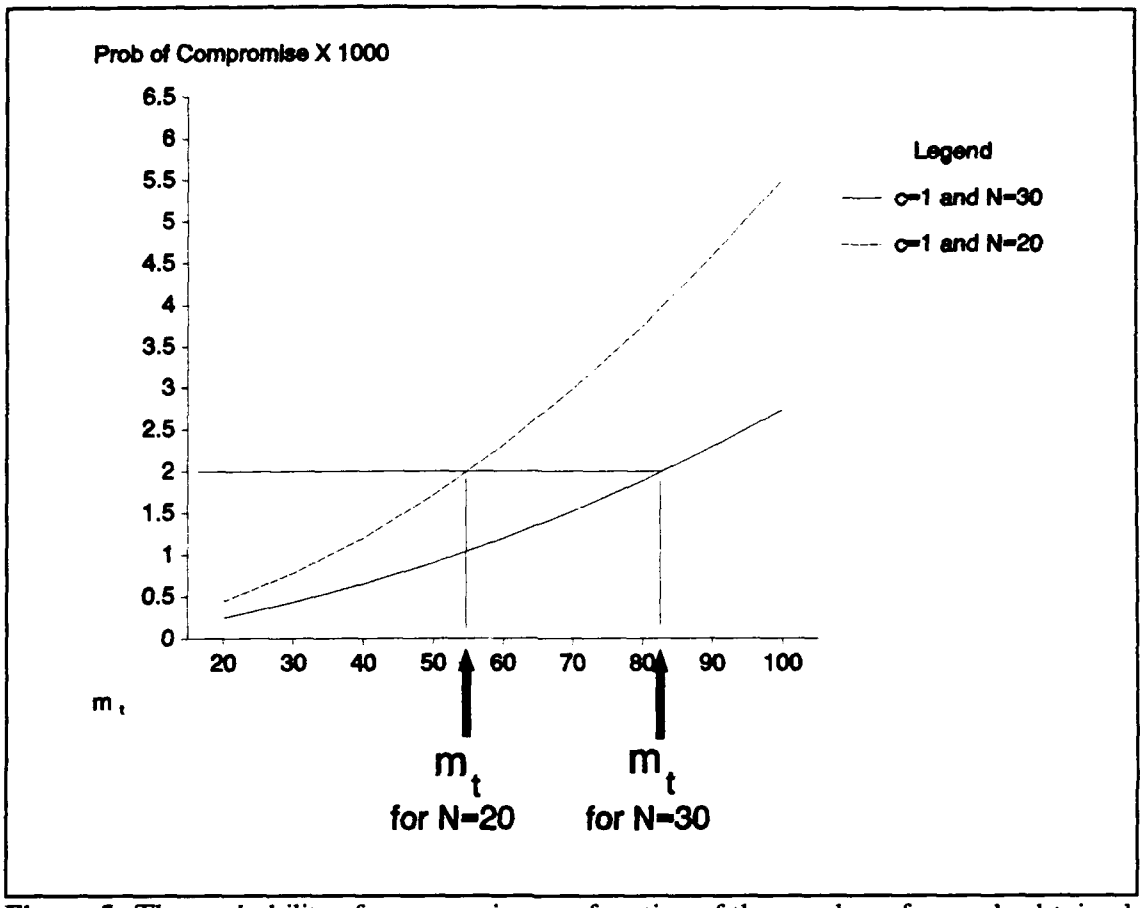

Figure 5. The probability of compromise as a function of the number of records obtained.

compromise as in Figure 5. Picking an acceptable limit on P_{comp} will provide m_t, a limit on m. Then it is only necessary to monitor, (perhaps via an Intrusion Detection System), the total number of records a user accesses from any particular class of N datasets to recognize a potential compromise by aggregation. The probability of this method failing to prevent a compromise would be the P_{comp} limit.

4.3 Recovering From an Aggregation Compromise

If the detection process fails and security is compromised, an investigation should be conducted to establish the cause so that recurrences are avoided. This requires incorporating a Learning Element into the DAT. The tool should make all aspects of the aggregation model easily available for inspection and manipulation.

For instance, suppose a compromise succeeded because a user was able to isolate records from the target database, enabling him to obtain a key subset with a total number of records smaller than m_t. This might indicate that a higher cue factor, (c), can be obtained than was originally believed. To enable the investigator to determine the actual cue factor the DAT should provide an environment which facilitates the investigation of such possibilities. The capability to profile a selection of records and to develop statistics of associations between these records and various key subsets in target classes would be useful. Doing so would help to identify the most likely approach taken by the user to compromise security.

The recovery mechanism must also, of course, allow the SSO to change various aspects of the DAT environment or the database as necessary to react to compromises. Setting new values for parameters such as T, N, and c should be possible. In addition, the ability to classify various key subsets in the data should be available.

4.4 Summarizing the Concept

We have developed a conceptual design of a DAT to address the aggregation problem that would involve the following:

- A database analysis and design tool: The focus of the *design* aspect of the tool would be to provide an architecture which optimizes the database for protection against aggregation. In this respect the tool would aid the database administrator in designing data associations and in structuring the database. The *analysis* aspect of the tool would be used to determine various parameters of the data such as the cue factor (c) and the P_{aimL} for various subsets. From this parametric information certain constraints such as m_t would be established.

- A database usage monitoring tool: The usage monitoring aspect of the DAT would be similar in concept to an Intrusion Detection System (IDS). In fact, the final implementation of the DAT should integrate its usage monitoring tool, if possible, with the available IDS technology. The objective of this aspect of the DAT would be to monitor database usage for various sets of N systems for which m_t constraints have been established.

- A Learning Tool: The learning aspect of the DAT would utilize the database analysis capability mentioned above along with various statistical or machine learning techniques to investigate compromises. The objective would be to discover what aspect of the current constraint parameter set was violated and determine if any parameters should be modified. It must be recognized that some compromises can be expected from correlation with outside information. Changing the constraint parameter set as a reaction to such violations may unnecessarily encumber the database while not significantly improving security.

All three aspects of the DAT should be integrated so that databases of security information can be shared. They would all need to be classified system high. Careful attention to the user interfaces would be a requirement to ensure that the various abstract concepts are represented as clearly as possible.

5.0 DIRECTIONS FOR FUTURE WORK

Further development of the aggregation concept should involve mapping the jigsaw model of aggregation to the IWSDB world. Work could then proceed to develop processes to implement the concepts identified. This involves the creation of a test database to validate the concepts and processes developed. Below we present a brief outline showing the directions for future work.

5.1 Mapping the Jigsaw Model to the IWSDB World

Mapping the jigsaw model to the IWSDB would begin with acquiring application specific knowledge. Knowing some details about the application is necessary to keep the proper focus, allowing us to protect sensitive data while not encumbering non-sensitive data.

Application specific knowledge can be acquired in the following ways:

- Reviewing the schema for the IWSDB
- Talking to SSO's or weapons systems program managers
    - to learn about the relationships between individual data items and potential aggregates
    - to learn about the types and frequency of queries likely to be enacted on the IWSDB

Once acquired, the application specific knowledge can be used to build data association models that would be implemented in an OODBMS to prevent aggregation. Figures 6 and 7 show examples of data associations that might be established for prevention and detection.

The mapping would then proceed by answering various questions such as:

- How are key parameters such as L, m, N, and n defined with respect to the IWSDB?
- How does the key subset concept carry over to the IWSDB world?
- Are derived parameters such as c, P_L, P_{simL}, and P_{mL} still valid as given in the jigsaw model, or do they require modification?

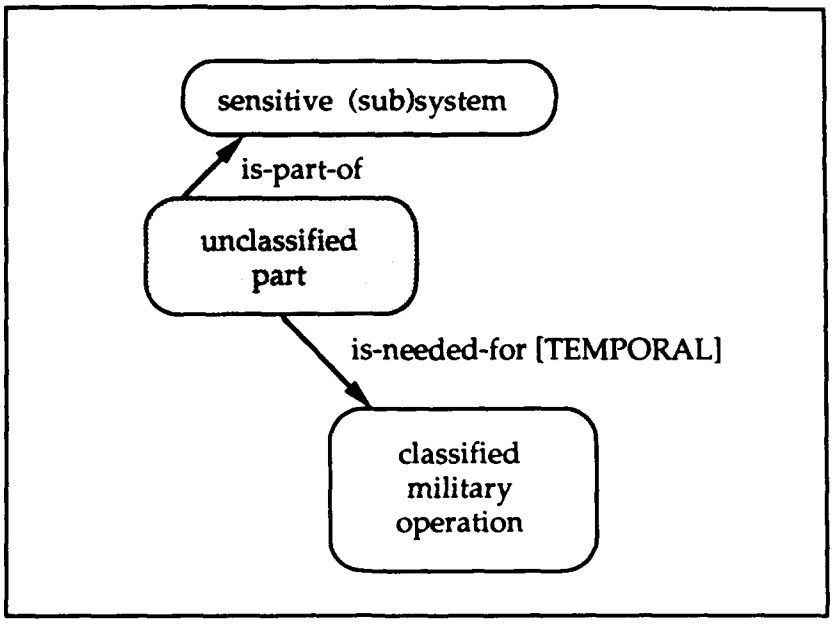

Figure 6. Sample Data Associations for Prevention

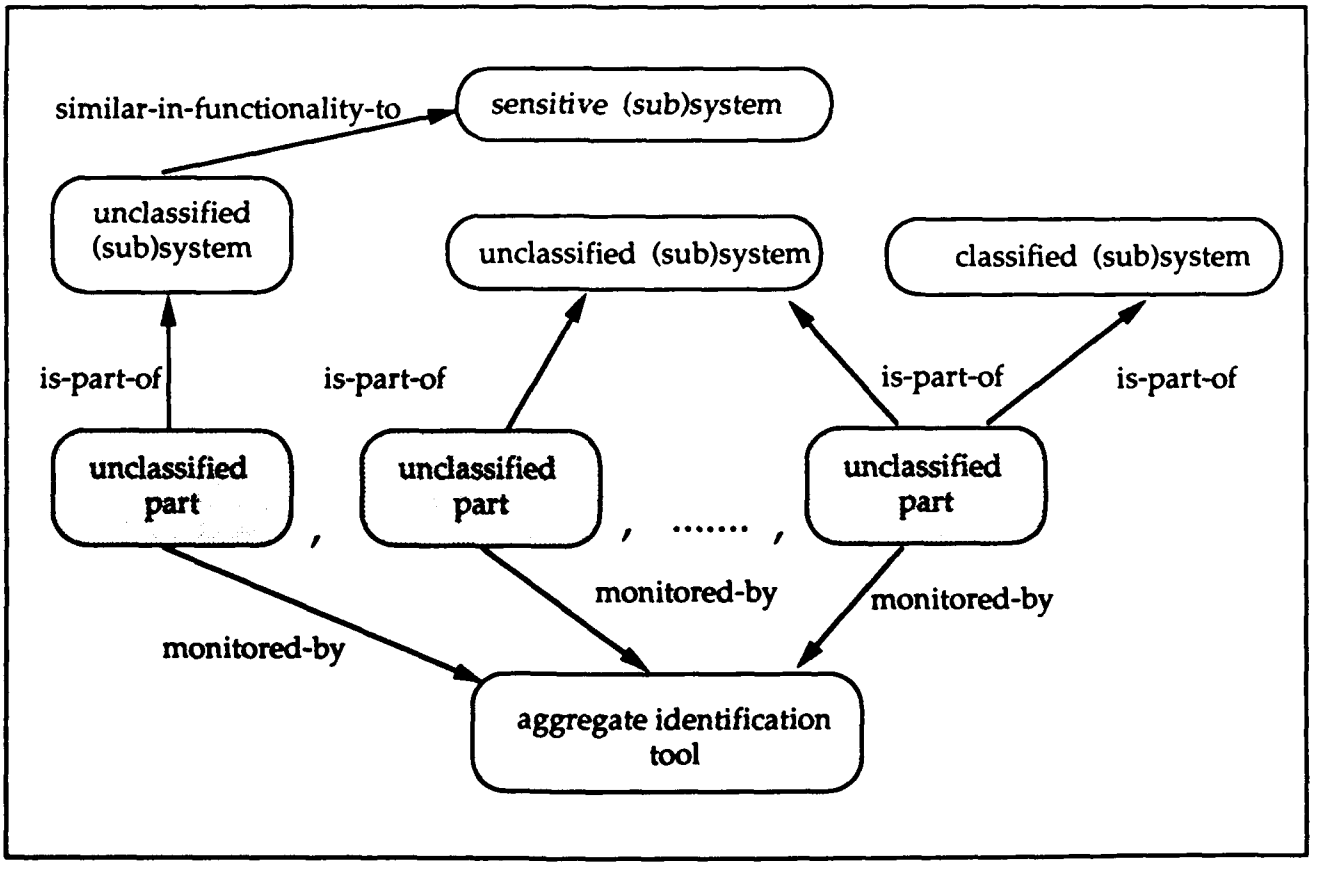

Figure 7. Detection Semantics

5.2 Developing Processes to Implement the Concepts Identified

Developing data design and analysis processes is probably the singly most important implementation activity, since succeeding activities draw heavily on these processes. Various questions have to be answered, such as the role commercial Object-Oriented Database Management Systems should play in supporting the implementation.

Object-oriented databases first became commercially available in 1987 when G-Base from Graphael was introduced. Since then, many other object-oriented database products have been developed. Besides the authentic object-oriented database products, some relational systems are evolving toward object orientation incrementally. While one cannot expect a pure object-oriented database system to become a de-facto standard in the near future, it is clear that the database industry is moving in an object-oriented direction. An area for further research is an investigation of the potential for a commercial OO database management system to provide support for our new data association models. Important considerations in evaluating such a system include:

- The capability to accommodate very long multimedia fields. (The IWSDB will likely contain image data such as engineering drawings.)

- An object-oriented extension to a query language. (Certain database users will be cleared to retrieve classified or sensitive aggregates. An OO extension will make such queries simpler.)

- Versioning, that is, access to previous states or alternate states of objects. (Some data may have temporal sensitivity as during a conflict or war and may require multiple versions of objects to accommodate changing constraints.)

- Security: (An OODBMS must also incorporate security primitives for accessing and updating objects.)

- Performance issues: (The IWSDB is likely to be both large and complex. An OODBMS that permits users to control the storage management is likely to perform better.)

Once questions addressing data design and analysis are answered, processes must be defined and implemented for establishing derived parameters such as the cue factor (c) and the identity of key subsets.

Identifying key subsets will probably remain a manual process. The personnel most likely to be able to identify potential aggregates are System Security Officers (SSOs) and program managers. In this task we would define and implement a well structured Monte Carlo experiment, as discussed in Section 4.1, using actual subsets of IWSDB data to identify

potential aggregates. The purpose of the experiment is to begin to identify the statistics of key subsets that would characterize a sensitive aggregate.

Determining (c) may require utilizing some formal technique to account for a variety of cues which may be available. One approach would be to use the Odds Likelihood formulation. The parameter c is defined as N times the probability of picking a target puzzle piece from among the aggregate of all pieces. If no information is available to discriminate the target pieces from non-target pieces, one would expect the probability to be that due to pure chance, 1/N, and c is then equal to 1. If some pieces of evidence exist which influence the decision in a positive sense, the probability will be greater than that expected from pure chance and c will be greater than 1. Deceptive evidence will make the probability less than that due to pure chance, making c less than 1.

An odds likelihood formulation provides a mechanism to update c from its a priori value of 1 based upon the various evidences available. To introduce this approach, note that the probability of an event can be expressed by its odds and vice versa as shown in equations 10 and 11.

$$O(x) = \frac{P(x)}{1 - P(x)} \tag{10}$$

$$P(x) = \frac{O(x)}{1 + O(x)} \tag{11}$$

In addition, as illustrated in equation 12, it may be shown that the odds of a hypothesis given a particular piece of evidence is related to the a priori odds of that hypothesis by the ratio of the probability of the evidence given the hypothesis is true to the probability of the evidence given that the hypothesis is false. This ratio, represented as lambda, is known as the odds likelihood ratio.

$$O(H \mid E) = \frac{P(E \mid H)}{P(E \mid \neg H)} O(H) = \lambda O(H) \tag{12}$$

Using this formulation, each characteristic of a puzzle piece which is in some way correlated with the piece's membership in the target puzzle can be assigned a likelihood ratio, lambda. Then the odds of picking a target piece by following the guidance provided by all such characteristics can be estimated as follows:

$$O(\textit{Membership} \mid E_1, E_2, \ldots E_f) = \lambda_1 \lambda_2 \ldots \lambda_f \, O(\textit{Membership}) \qquad (13)$$

Converting back to a probability formulation and solving for c, as represented in equations 14 and 15, leads to an expression for c in terms of the likelihood ratio products.

$$P(x) = \frac{c}{N} = \frac{O(x)}{O(x)+1} \rightarrow c = N \frac{O}{1+O} \qquad (14)$$

$$c = N \frac{\lambda_1 \lambda_2 \ldots \lambda_F}{\lambda_1 \lambda_2 \ldots \lambda_F + N-1} \qquad (15)$$

This outlines the approach for determining c via an odds likelihood formulation. Implementing this approach will require considerable experience to identify the various characteristics and to establish their likelihood ratios.

APPENDIX A

BIBLIOGRAPHY

This bibliography was compiled by IIT Research Institute (IITRI) for the Defense Logistics Agency (DLA) in support of the Computer Aided Acquisition and Logistics Support (CALS) Database Usage Analysis Tool Study. The purpose of this study is to develop the specifications for a tool which will monitor CALS database usage in real-time to prevent unauthorized access to potentially sensitive and proprietary data. While not classified at the data element level, certain data, when aggregated, may be sensitive or even classified. IITRI is working on defining a concept for a tool that will help to minimize this risk; that is, minimize the possibility of a user inferring high-level information based upon lower level visible data.

This is the second and final update to the bibliography which was originally published in January 1991. The citations appear in ascending order by publication date. Some of the citations on the following pages were identified by reviewing holdings in Rome Laboratory's (previously, Rome Air Development Center) technical library and searching the INSPEC database (the Database for Physics, Electronics and Computing database) on the DIALOG computer database service. INSPEC corresponds to the three Science Abstracts print publications: *Physics Abstracts, Electrical and Electronics Abstracts,* and *Computer and Control Abstracts.* Approximately twenty-five citations were derived from documents obtained directly from SRI.

Requirements and Model for IDES - A Real-time Intrusion-Detection Expert System

Space and Naval Warfare Command (SPAWAR) Final Report

Author(s)	Denning, D.; Neumann, P.
Author Affiliation	SRI International
Citation Source	Report Documentation Page; DD Form 1473
Document Availability	SPAWAR 83F830100 Lt. Commander Phil Myers SPAWAR 814T; 202/692-8484
Document Source	Space and Naval Warfare Command (SPAWAR) Final Report
Publication Date	August 1985
Publisher	Space and Naval Warfare Command
Subject Treatment	Practical
Document Type	Final Report
References	8 Additional References

Abstract This report describes the basis for IDES, an Intrusion-Detection Expert System that aims to detect intrusions, penetrations, and other forms of computer abuse while they are in progress by looking for abnormal patterns of system use.

Descriptors IDES, intrusion detection, security, expert systems

An Intrusion-Detection Model

Proceedings of 1986 Symposium on Security and Privacy

Author(s)	Denning, D.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: 1986 Symposium on Security and Privacy
Publication Date	April 1986
Publisher	IEEE Computer Society
Subject Treatment	Practical
Document Type	Conference Paper
References	3 Additional References

Abstract A model of a real-time intrusion-detection expert system capable of detecting break-ins, penetrations, and other forms of computer abuse is described. The model is based on the hypothesis that security violations can be detected by monitoring a system's audit records for abnormal patterns of system usage. The model includes profiles for representing the behavior of subjects with respect to objects in terms of metrics and statistical models, and rules for acquiring knowledge about this behavior from audit records and for detecting anomalous behavior. The model is independent of any particular system, application environment, system vulnerability, or type of intrusion, thereby providing a framework for a general-purpose intrusion-detection expert system.

Descriptors intrusion detection, expert system, anomalous behavior

Semantic Database Modeling: Survey, Applications, and Research Issues

Author(s)	Hull, Richard and King, Roger
Author Affiliation	University of Southern CA, Los Angeles, CA, and University of Colorado, Boulder, CO
Citation Source	Review of Document
Document Availability	ACM Computing Surveys
Document Source	Association for Computing Machinery
Publication Date	1987
Publisher	Association for Computing Machinery
Subject Treatment	Practical
Document Type	Journal Article
References	126 Additional References

Abstract Most common database management systems represent information in a simple record-based format. Semantic modeling provides richer data structuring capabilities for database applications. In particular, research in this area has articulated a number of constructs that provide mechanisms for representing structurally complex interrelations among data typically arising in commercial applications. In general terms, semantic modeling complements work on knowledge representation in (artificial intelligence) and on the new generation of database models based on the object-oriented paradigm of programming languages.

This paper presents an in-depth discussion of semantic data modeling. It reviews the philosophical motivations of semantic models including the need for high-level modeling abstractions and the reduction of semantic overloading of data type constructors. It then provides a tutorial introduction to the primary components of semantic models which are the explicit representation of objects, attributes of, and relationships among objects, type constructors for building complex types, ISA relationships, and derived schema components. Next, a survey of the prominent semantic models in the literature is presented. Further, since a broad area of research has developed around semantic modeling, a number of related topics based on these models are discussed including data languages, graphical interfaces, theoretical investigations, and physical implementation strategies.

Descriptors conceptual database design, entity-relationship model, functional data model, knowledge representation, semantic database model

The Elements of Artificial Intelligence
An Introduction Using LISP

Author(s)	Tanimoto, Steven L.
Author Affiliation	University of Washington, Seattle, WA
Citation Source	Review of Document
Document Availability	Computer Science Press
Document Source	Computer Science Press
Publication Date	1987
Publisher	Computer Science Press, Rockville, MD
Subject Treatment	Theoretical and Practical
Document Type	Book
References	7 Additional References

Abstract The rapidly expanding subject of Artificial Intelligence requires professionals who have a firm grasp of both its scientific principles and its implementation techniques. Without the principles, the practitioner flounders whenever he reaches the limits of his tools. Without implementation experience, one has very limited intuition about what is feasible and how a new project should be organized. This book presents both the principles and the chief programming techniques of Artificial Intelligence. The table of contents follows by chapter number:

1 Introduction to AI
2 Programming in LISP
3 Productions and Matching
4 Knowledge Representation
5 Search
6 Logical Reasoning
7 Probabilistic Reasoning
8 Learning
9 Natural Language Understanding
10 Vision
11 Expert Systems
12 The Future

Descriptors artificial intelligence, LISP, AI programming, knowledge representation, inference, fuzzy logic, inference networks, learning, natural language understanding, machine vision, expert systems

A Prototype Real-Time Intrusion-Detection Expert System

Proceedings of the 1988 IEEE Symposium on Security and Privacy

Author(s)	Lunt, T.; Jagannathan, R.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: 1988 IEEE Symposium on Security and Privacy
Publication Date	April 1988
Publisher	IEEE
Subject Treatment	Practical
Document Type	Conference Paper
References	7 Additional References

Abstract This paper describes the design and implementation of a prototype intrusion-detection expert system (IDES) developed at SRI International. IDES is based on the concept that an intrusion manifests itself as a departure from expected behavior for a user. The prototype monitors users on a remote system using audit records which characterize their activities. It adaptively learns normal behavior of each user and detects and reports anomalous user behavior in real-time.

Descriptors intrusion detection, expert system, IDES, security, anomalous behavior

Inference Aggregation Detection In Database Management Systems

Proceedings of the 1988 IEEE Symposium on Security and Privacy

Author(s)	Hinke, Thomas H.
Author Affiliation	TRW Defense Systems Group Redondo Beach, CA
Citation Source	Review of Document
Document Availability	TRW Defense Systems Group Redondo Beach, CA
Document Source	Proceedings: 1988 IEEE Symposium on Security and Privacy
Publication Date	April 1988
Publisher	IEEE Computer Society Press
Subject Treatment	Practical
Document Type	IEEE Journal Article
References	6 Additional References

Abstract This paper makes seven contributions to security aggregation research. It identifies inference aggregation and cardinality aggregation as two distinct aspects of the aggregation problem. The paper develops the concept of a semantic relationship graph to describe the relationships between data and then presents inference aggregation as the problem of finding alternative paths between vertices on the graph. An algorithm is presented for processing the semantic relationship graph to discover whether potential inference aggregation problems exist. A method of detecting some aggregation conditions within the DBMS is presented which uses the normal DBMS query language and adds additional catalytic data to the DBMS to permit a query to make the inference. The paper also suggests use of set theory to describe aggregation conditions and the addition of set operations to the DBMS to permit the description of aggregation detection queries.

Descriptors aggregation, semantic modeling, inference aggregation, cardinality aggregation

NIDX - An Expert System for Real-Time Network Intrusion Detection

Proceedings of theComputer Networking Symposium

Author(s)	Bauer, David S.; Koblentz, Michael E.
Author Affiliation	Bell Communications Research, Inc. Piscataway, NJ
Citation Source	Review of Document
Document Availability	IEEE Computer Society
Document Source	Proceedings: Computer Networking Symposium
Publication Year	April 1988
Publisher	IEEE Computer Society Press, Washington, DC
Subject Treatment	Practical
Document Type	Conference Paper
References	9 Additional References

Abstract A knowledge-based prototype Network Intrusion Detection Expert System (NIDX) for the Unix system V environment is described. NIDX combines knowledge describing the target system, history profiles of users' past activities, and intrusion detection heuristics forming a knowledge-based system capable of detecting specific violations that occur on the target system. Intrusions are detected by classifying user activity from a real-time audit trail of UNIX system calls; then, using system-specific knowledge and heuristics about typical intrusions and attack techniques, determines whether or not the activity is an intrusion. This paper describes the NIDX knowledge base, UNIX system audit trail mechanism and history profiles, and demonstrates the knowledge-based intrusion detection process.

Descriptors knowledge-based intrusion detection, real-time network intrusion detection, user profile analysis, UNIX audit trail analysis

Secure Distributed Data Views

Vol. 1: Security Policy and Policy Interpretation for a Class A1 Multilevel Secure Relational Database System

Author(s)	Lunt, T.; Neumann, P.; Denning, D.; and Schell, R.; Heckman, M.; Shockley, W .
Author Affiliation	SRI International, Menlo Park, CA Gemini Computers, Inc., Monterey, CA
Citation Source	Review of Document
Document Availability	Not Known; Work Performed for Rome Air Development Center Under Contract F30602-85-C-0243
Document Source	SRI International, Menlo Park, CA
Publication Date	August 1988
Publisher	Not Known
Subject Treatment	Practical
Document Type	Technical Report
References	33 Additional References

Abstract This report describes a security policy for a secure relational database system. This policy is intended to meet the security policy requirement specified in the *DoD Trusted Computer System Evaluation Criteria*. Because the policy is intended for a relational database management system, it goes beyond policies that the reader may be familiar with for general-purpose systems. However, it also addresses the requirements considered applicable to general-purpose systems and can serve as a useful guide to those who are called upon to produce a policy statement that will satisfy the *Criteria*. The development of a security policy is the first task of a three-year project to design a multilevel secure database system that will satisfy the criteria for Class A1.

Descriptors Security policy, trusted systems, multilevel security, computer security, database management system, secure database management system, database security, security kernel, reference monitor, relational database

Automated Audit Trail Analysis and Intrusion Detection: A Survey

Proceedings of the11th National Computer Security Conference

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	SRI International, Menlo Park, CA
Publication Date	October 1988
Publisher	Not Known
Subject Treatment	Practical
Document Type	Conference Paper
References	23 Additional References

Abstract Today's computer systems are vulnerable to both abuse by insiders and penetration by outsiders, as evidenced by the growing number of incidents reported in the press. Because closing all security loopholes from today's systems is infeasible, and since no combination of technologies can prevent legitimate users from abusing their authority in a system, auditing is viewed as the last line of defense. What is needed are automated tools to analyze the vast amount of audit data for suspicious user behavior. This paper presents a survey of the automated audit trail analysis techniques and intrusion-detection systems that have emerged in the past several years.

Descriptors intrusion detection, audit trail analysis

IDES: The Enhanced Prototype
A Real-Time Intrusion-Detection Expert System

Author(s)	Lunt, T.; Jagannathan, R.; Lee, R.; Listgarten, S.; Edwards, D.; Neumann, P.; Javitz, H.; Valdes, A.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	SRI International, Menlo Park, CA
Publication Date	October 1988
Publisher	SRI International, Menlo Park, CA
Subject Treatment	Practical
Document Type	Technical Report
References	8 Additional References

Abstract This report describes the design and implementation of a real-time intrusion-detection expert system (IDES) designed and developed by SRI International. IDES is an independent system that monitors the activities of different types of subjects, such as users and remote hosts of a target system, to detect security violations by both insiders and outsiders as they occur. IDES adaptively learns subjects' behavior patterns over time and detects behavior that deviates from these patterns. IDES also has an expert system component that can be used to encode information about known system vulnerabilities and intrusion scenarios.

Descriptors intrusion detection, expert system, IDES

Haystack: An Intrusion Detection System

Proceedings of the Fourth Aerospace Computer Security Applications Conference

Author(s)	Smaha, Stephen E.
Author Affiliation	Tracor Applied Sciences, Inc., Austin, TX
Citation Source	Review of Document
Document Availability	IEEE Computer Society
Document Source	Proceedings: Fourth Aerospace Computer Security Applications Conference
Publication Date	December 1988
Publisher	IEEE Computer Society Press, Washington, DC
Subject Treatment	Practical
Document Type	Conference Paper
References	4 Additional References

Abstract Haystack is a prototype system for the detection of intrusions in multi-user Air Force computer systems. Haystack reduces voluminous system audit trails to short summaries of user behaviors, anomalous events, and security incidents. This is designed to help the System Security Officer (SSO) detect and investigate intrusions, particularly by insiders (authorized users). Haystack's operation is based on behavioral constraints imposed by security policies and on models of typical behavior for user groups and individual users.

Descriptors intrusion detection, anomaly detection, behavior analysis

Secure Distributed Data Views: Identification of Deficiencies and Directions for Future Research

Final Report, Volume 4

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	Not Known; Work Performed for Rome Air Development Center Under Contract F30602-85-C-0243
Document Source	SRI International, Menlo Park
Publication Date	31 January 1989
Publisher	Not Known
Subject Treatment	Theoretical and Practical
Document Type	Final Report
References	72 Additional References

Abstract SeaView was a three-year project that was a pioneer in designing a multilevel relational database system that meets the *Criteria* for Class A1. SeaView significantly advanced the state of the art in database security. This report discusses further research that could be done to extend SeaView's ideas. The areas discussed include aggregation and inference, concurrency, distributed data, discretionary security, a query language for multilevel data, classification constraints, and object-oriented database systems.

Descriptors aggregation, inference, concurrency, distributed data, discretionary security, multilevel data, classification constraints

Real-Time Intrusion Detection

Proceedings : COMPCON Spring '89

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: COMPCON Spring '89
Publication Date	27 February 1989
Publisher	Not Known
Subject Treatment	Practical
Document Type	Conference Paper
References	6 Additional References

Abstract This paper describes a real-time intrusion-detection expert system (IDES) that observes user behavior on a monitored computer system and adaptively learns what is normal for individual users, groups, remote hosts, and the overall system behavior. Observed behavior is flagged as a potential intrusion if it deviates significantly from the expected behavior or if it triggers a rule in the expert-system rule base.

Descriptors IDES, intrusion detection, expert system

Knowledge-Based Intrusion Detection

Proceedings of the Annual AI Systems in Government Conference

Author(s)	Lunt, T.F.; Jagannathan, R.; Lee R.; Whitehurst, A.; and Listgarten, S.
Author Affiliation	SRI International, Menlo Park, CA Stanford University, Stanford, CA
Citation Source	INSPEC; DIALOG File #13
Document Availability	IEEE Catalog No. 89CH2715-1
Document Source	Proceedings: Annual AI Systems in Government Conference
Publication Year	1989
Publisher	IEEE Computer Society Press, Washington, DC
Subject Treatment	Practical
Document Type	Conference Paper
References	10 Additional References

Abstract The authors describe the expert-system aspects of IDES (Intrusion-Detection Expert System). A system for computer intrusion detection, IDES uses two distinct approaches to detect anomalies (which could signify intrusions) in a computer system, namely, statistical and rule-based anomaly detection. In the statistical approach, recent behavior of a subject of a computer system is compared with observed behavior and any significant deviation is considered anomalous. In the rule-based approach, acceptable behavior of a subject is captured by a set of rules which is used to identify anomalous observed behavior. The authors claim that integrating the two approaches in IDES provides for a comprehensive system for detecting intrusions as they occur.

Descriptors auditing, expert systems, statistical anomaly detection, statistical intrusion detection, automated audit trail analysis, intrusion-detection expert system, rule-based anomaly detection

Aggregation and Inference: Facts and Fallacies

Proceedings of the 1989 IEEE Symposium on Research in Security and Privacy

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	SRI International, Menlo Park
Publication Date	May 1989
Publisher	Not Known
Subject Treatment	Practical
Document Type	Conference Paper
References	16 Additional References

Abstract This paper examines inference and aggregation problems that can arise in multilevel relational database systems and points out some fallacies in our thinking about these problems that may hinder real progress from being made toward their solution. Although others have done some initial research toward solving inference problems, aggregation has been treated only superficially in the literature. This paper attempts to lay a firmer foundation for a theory of these problems. Several types of problems are identified and approaches toward their solution suggested.

Descriptors aggregation, inference, multilevel relational database systems

Deducibility Security with Dynamic Level Assignments

Proceedings of the Computer Security Foundations Workshop II

Author(s)	Sutherland, I.; Perlo, S.; Varadarajan, R.
Author Affiliation	Odyssey Research Associates, Inc., Ithaca, NY
Citation Source	Review of Document
Document Availability	IEEE Computer Society
Document Source	Proceedings: Computer Security Foundations Workshop II
Publication Year	1989
Publisher	IEEE Computer Society Press, Washington, DC
Subject Treatment	Theoretical
Document Type	Conference Paper
References	2 Additional References

Abstract The authors give a generalization of the definition of security for state machines given by D. Sutherland (Proceedings of the 9th National Computer Security Conference, September 1986). The generalization allows the security levels of inputs and outputs to be assigned dynamically. Its aim is merely to say what it means to infer high-level information from low-level information when the definitions of what is high and low can change. Although the generalization supports the modeling of things like login and reclassification, it does not give any guidance about how to do login or reclassification correctly. It merely allows such procedures to be represented; this cannot be done in a straightforward way with previous models.

Descriptors deducibility security model, dynamic security assignments, security for state machines

Multilevel Security for Knowledge Based Systems
Final Report

Author(s)	Lunt, Teresa F.; Garvey, Thomas D.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	Not Known; Work Performed for Rome Air Development Center Under Contract F30602-87-D-0094
Document Source	IIT Research Institute, Lanham, MD
Publication Year	14 August 1989
Publisher	Not Known
Subject Treatment	Practical
Document Type	Final Report
References	20 Additional References

Abstract Work aimed at defining a multilevel, mandatory security policy for knowledge-based systems is discussed. Two distinct issues are addressed: an effective implementation formalism based on a multilevel, object-oriented programming paradigm, and requirements for ensuring the correctness of handling multilevel objects within a single access class are defined, and a method by which multilevel objects may be used to implement a simple knowledge-based system built on production rules is outlined. The argument is made that the issues regarding correctness are similar to those of truth maintenance in standard knowledge-based systems and may be addressed by similar methods.

Descriptors multilevel security, inference, knowledge-based system, truth maintenance

Secure Knowledge-Based Systems

Interim Technical Report

Author(s)	Lunt, Teresa F. and Millen, Jonathan K.
Author Affiliation	SRI International, Menlo Park, CA, and The MITRE Corporation, Bedford, MA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	IIT Research Institute, Lanham, MD
Publication Date	29 August 1989
Publisher	SRI International, Menlo Park, CA
Subject Treatment	Practical
Document Type	Interim Technical Report
References	18 Additional References

Abstract This report proposes a security model and suggests a design strategy for knowledge-based systems that is based on the object-oriented model of data. The object oriented programming paradigm is a natural medium in which to implement a knowledge-based system. It supports the notion of a class hierarchy, an essential ingredient of knowledge-based systems. Unlike the relational model, it naturally captures the semantics of the information it contains. This report first discusses the essential features of a general object system model, and then extends the object model to incorporate mandatory label-based security. The report goes to show how typical database security and integrity policies can be supported by this model with special attention to inference problems and integrity constraints.

Descriptors knowledge based systems, multilevel security, object-oriented, mandatory security, inference

Overview of Security Technology Efforts at Bell Communications Research

Proceedings of the 1989 International Carnahan Conference on Security Technology

Author(s)	Schwartz, Barry K.
Author Affiliation	Bellcore, Morristown, NJ
Citation Source	Review of Document
Document Availability	IEEE Catalog No. 89CH2774-8
Document Source	Proceedings: 1989 International Carnahan Conference on Security Technology (pp. 79-81)
Publication Year	1989
Conference Information	3-5 October 1989; Zurich, Switzerland
Publisher	ETH Zentrum-KT, Zurich, Switzerland
Subject Treatment	Practical
Document Type	Conference Paper
References	No Additional References

Abstract This paper describes a four point security technology plan that has been implemented at Bellcore to improve the security of both its telecommunications network and telecommunications operations in the United States. The four points of the plan are: 1) for all existing systems, ensure that we appropriately use existing security features, 2) carefully review existing systems for security vulnerabilities, and fix known holes, 3) for new systems, architect security in from the very beginning, and 4) actively develop or seek out, and deploy new technologies in order to improve security and stay ahead of the adversary.

Descriptors biometric authentication, encryption, expert systems technology

Intrusion Detection: An Application of Expert Systems to Computer Security

Proceedings of the 1989 International Carnahan Conference on Security Technology

Author(s)	Bauer, D.S.; Eichelman, F.R., II; Herrera, R.M.; Irgon, A.E.
Author Affiliation	Bellcore, Piscataway, NJ
Citation Source	INSPEC; DIALOG File #13
Document Availability	IEEE Catalog No. 89CH2774-8
Document Source	Proceedings: 1989 International Carnahan Conference on Security Technology (pp. 97-100)
Publication Year	1989
Publisher	ETH Zentrum-KT, Zurich, Switzerland
Subject Treatment	Practical
Document Type	Conference Paper
References	5 Additional References

Abstract Intrusion detection is an area of computer security that focuses on developing the technology to detect intruders on computer systems in near real time through the use of software systems that automatically analyze computer system audit trails. An overview of current intrusion detection research and technology is presented. The Network Intrusion Detection Expert System (NIDX) is described as an example of an intrusion detection system. Its system architecture, detection principles, and detection strategy are discussed.

Descriptors expert systems, intrusion detection, computer system audit trails, Network Intrusion Detection Expert System, NIDX, detection principles

Foiling the Wiley Hacker: More than Analysis and Containment

Author(s)	Kluepfel, Henry M.
Author Affiliation	Bellcore, Morristown, NJ
Citation Source	Review of Document
Document Availability	Unknown
Document Source	International Carnahan Conference on Security Technology
Publication Date	1989
Publisher	ICCST, Zurich, Switzerland
Subject Treatment	Practical
Document Type	Conference Paper
References	16 Additional References

Abstract This paper looks at the methods and tools used by system intruders. It analyzes the development of the hacker, his motivation, his environment, and the burglar tools used for system intrusion. It also probes the nature of the vulnerable networking environments that are the target of intrusions. More important, it will address turning the tables on these intruders with their own tools and techniques. Besides reacting to attacks, there are many opportunities to learn from the intruders and design that knowledge into new defensive solutions for securing computer-based systems. Having presented the problem, the paper presents a strategy to defend and thwart such intrusions in our increasingly networked and distributed computing and telecommunications environments.

Descriptors intrusion detection, network security

Detection of Anomalous Computer Session Activity

Proceedings of the 1989 IEEE Symposium on Security and Privacy

Author(s)	Vaccaro, H.S. and Liepins, G.E.
Author Affiliation	Los Alamos National Laboratory, Los Alamos, NM and Oak Ridge National Laboratory, Oak Ridge, TN
Citation Source	Review of Document
Document Availability	IEEE Computer Society
Document Source	Proceedings: 1989 IEEE Computer Society Symposium on Security and Privacy
Publication Date	1989
Publisher	IEEE Computer Society Press, Washington, DC
Subject Treatment	Practical
Document Type	Conference Paper
References	9 Additional References

Abstract This paper briefly discusses Wisdom and Sense (W&S), a computer security anomaly detection system developed at Los Alamos National Laboratory (LANL). Anomaly detection provides another layer of defense against computer misuse after physical security and access security. W&S is statistically based. It automatically generates rules from historical data and in terms of those rules, identifies computer transactions that are at variance with historically established usage patterns. Issues addressed in this paper include how W&S generates rules from a necessarily small sample of all possible transactions, how W&S deals with inherently categorical data, and how W&S assists system security officers in their review of audit logs.

Descriptors anomaly detection, computer security, usage analysis

Database Inference Controller

Rome Air Development Center Final Technical Report

Author(s)	Buczkowski, Leon J.; Perry, E. L.; Lee, David H.
Author Affiliation	Ford Aerospace Corporation Colorado Springs, CO
Citation Source	Report Documentation Page; DD Form 1473
Document Availability	RADC-TR-89-329 Mr. Joseph V. Giordano RADC/COTD; 315/330-2925
Document Source	Rome Air Development Center Final Technical Report No. RADC-TR-89-329
Publication Date	January 1990
Publisher	Rome Air Development Center
Subject Treatment	Practical
Document Type	Final Technical Report
References	23 Additional References

Abstract The primary objective of this program was an investigation and subsequent design of a Database Inference Controller (DBIC), a knowledge-based tool or set of tools, used off-line to detect and correct logical inferences in multilevel secure (MLS) databases. The program involved developing a realistic working example of an MLS database; evaluating methods for identifying, modeling, and quantifying inference in MLS databases; and incorporating the results from a top-level DBIC design. The DBIC design implements the technique of probabilistic knowledge modeling to identify inference, creating and using a probabilistic inference network integrated with a semantic model of the target database. The probabilistic inference network, derived from an elaboration of the Command and Control (C^2) system's security classification policy, is a structure that identifies the logical dependencies of classified parameters on aggregates of objects at lower classifications. The DBIC top-level design incorporates the results of the investigation into a knowledge-based architecture.

Descriptors database security inference, knowledge-based systems, multilevel security, semantic models, expert systems, probabilistic inference networks

Intelligent Database Systems

Rome Air Development Center Final Technical Report

Author(s)	Morgenstern, Matthew
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Report Documentation Page; DD Form 1473
Document Availability	RADC-TR-90-58 Mr. Joseph V. Giordano RADC/COTD; 315/330-2925
Document Source	Rome Air Development Center Final Technical Report No. RADC-TR-90-58
Publication Date	March 1990
Publisher	Rome Air Development Center
Subject Treatment	Practical
Document Type	Final Technical Report
References	74 Additional References

Abstract The objective of this project has been the design of a new generation of information system which is knowledgeable about the application it serves. The Intelligent Constraints, Active Data System (ICADS) which we have designed represents application knowledge and uses it both to ensure the consistency and reliability of the information, and to initiate active responses based upon the current status of the environment and needs of the users. Our approach in ICAD tightly integrates techniques from databases and from relevant Artificial Intelligence disciplines through our development of intelligent constraints to support active data objects. We formally define a constraint logic language, where a declarative constraint represents an assertion or invariant condition describing the application. The use of pattern-based specification makes possible the application of constraints to multiple data models and schemas. Each constraint provides a higher level specification than a rule, since a set of several condition-action rules supports each constraint. Our constraint language readily expresses semantic constructs commonly found in semantic-models. In addition, we provide enforcement for such semantics.

Descriptors intelligent database systems, active database systems, constraints, fault tolerance, multilevel security

Multilevel Security for Knowledge-Based Systems

Author(s)	Garvey, T.; Lunt, T.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	SRI International, Menlo Park, CA
Publication Date	4 May 1990
Publisher	SRI International, Menlo Park, CA
Subject Treatment	Practical
Document Type	Technical Report
References	18 Additional References

Abstract The authors discuss ongoing work aimed at defining a multilevel, mandatory security policy for knowledge-based systems. Two distinct issues are addressed: an effective implementation formalism based on a multilevel, object-oriented programming paradigm and requirements for ensuring the correctness of inferences computed on the basis of possibly contradictory information from different access classes. The authors define requirements for an object-oriented system capable of handling multilevel objects within a single access class. A method by which multilevel objects may be used to implement a simple knowledge-based system based on production rules is outlined. The authors present the argument that the issues regarding correctness are similar to those of truth-maintenance in standard knowledge-based systems and may be addressed by similar methods.

Descriptors Multilevel security, knowledge-based systems, multilevel objects, truth-maintenance

A Real-Time Intrusion-Detection Expert System

Technical Report

Author(s)	Lunt, T.; Tamaru, A.; Gilham, F.; Jagannathan, R.; Jalali, C.; Javitz, H.; Valdes, A.; Neumann, P.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	Not Known; Work Performed for U.S. Navy, SPAWAR Under Contract N00039-89-C-0050
Document Source	SRI International, Menlo Park, CA
Publication Date	June 1990
Publisher	Not Known
Subject Treatment	Practical
Document Type	Technical Report
References	17 Additional References

Abstract A real-time intrusion-detection expert system (IDES) has been designed and developed by SRI International. IDES is an independent system that observes user behavior on a monitored computer system and flags suspicious events. IDES monitors the activities of individual users, groups, remote hosts, and entire systems and detects suspected security violations by both insiders and outsiders as they occur. IDES adaptively learns users' behavior patterns over time and detects behavior that deviates from these patterns. IDES also has an expert system component that can be used to encode information about known system vulnerabilities and intrusion scenarios. Integrating the two approaches makes IDES a comprehensive system for detecting intrusions as well as misuse by authorized users. IDES has been enhanced to run under GLU, a language supporting distributed, parallel computation; GLU enhances flexibility and system fault tolerance.

Descriptors intrusion detection, expert system, IDES, distributed systems

The SeaView Security Model

Author(s)	Lunt, T.; Denning, D.; Schell, R.; Heckman, M; Shockley, W.
Author Affiliation	SRI International, Menlo Park, CA Digital Equipment Corp., Palo Alto, CA Gemini Computers Inc., Carmel, CA Digital Equipment Corp., Mountain View, CA
Citation Source	Review of Document
Document Availability	IEEE
Document Source	IEEE Transactions on Software Engineering, Vol. 16, No.6
Publication Date	June 1990
Publisher	IEEE Computer Society Press, Washington, DC
Subject Treatment	Practical
Document Type	Journal Article
References	23 Additional References

Abstract A multilevel database system is intended to provide the security needed for database systems that contain data at a variety of classifications and serve a set of users having different clearances. This paper describes a formal security model for such a system. The model is formulated in two layers, one corresponding to a reference monitor that enforces mandatory security, and the second, an extension of the standard relational model, defining multilevel relations and formalizing policies for labeling new and derived data, data consistency, and discretionary security. The model also defines application-independent properties for entity integrity, referential integrity, and polyinstantiation integrity.

Descriptors security model, multilevel secure database system, SeaView

Using Statistics to Track Intruders

Proceedings of the Joint Statistical Meetings of the American Statistical Association

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: Joint Statistical Meetings of the American Statistical Association
Publication Date	August 1990
Publisher	American Statistical Association
Subject Treatment	Practical
Document Type	Conference Paper
References	15 Additional References

Abstract This paper describes a real-time intrusion-detection expert system (IDES) that observes user behavior on a monitored computer system and adaptively learns what is normal for individual users, groups, remote hosts, and the overall system behavior. Observed behavior is flagged as a potential intrusion if it deviates significantly from the expected behavior or if it triggers a rule in the expert-system rule base.

Descriptors IDES, intrusion detection, expert system

Computer System Intrusion Detection

Final Technical Report

Author(s)	Hubbard, B.; Haley, T.; McAuliffe, N.; Schaefer, L.; Kelem, N.; Wolcott, D.; Feiertag, R.; Schaefer, M.
Author Affiliation	Trusted Information Systems, Inc. Mountain View, CA
Citation Source	Review of Document
Document Availability	Not Known; Work Performed for Rome Air Development Center Under Contract No. F30602-87-D-0093
Document Source	BBN Systems and Technologies Final Technical Report No. E002
Publication Date	20 September 1990
Publisher	BBN Systems and Technologies
Subject Treatment	Practical
Document Type	Final Technical Report
References	52 Additional References

Abstract A recent study of audit in Trusted Database Management Systems (TDBMS) environments has shown that there is value in collecting audit data from multiple levels of abstraction within the computer system, each level corresponding to a mode of interaction between the user and the system (e.g., the operating system level, the database management system level). A vast amount of audit data can be collected at each of these levels of abstraction. In addition, the audit data from all these levels needs to be correlated and analyzed. Therefore, there is a need for automated tools to aid in analyzing the audit data to look for suspicious user behavior and for unexpected system behavior. This work has continued that line of research by examining state-of-the-art in intrusion detection technology, identify issues which must be considered if the Air Force is to make effective use of intrusion detection technology, and make near and long term recommendations.

Descriptors multilevel security, audit data, intrusion detection, Trusted Database Management Systems (TDBMS)

IDES: An Intelligent System for Detecting Intruders

Proceedings of the Computer Security, Threat and Countermeasures Symposium

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: Computer Security, Threat and Countermeasures Symposium
Publication Date	November 1990
Publisher	Not Known
Subject Treatment	Practical
Document Type	Conference Paper
References	10 Additional References

Abstract This paper describes a real-time intrusion-detecftion expert system (IDES) that observes user behavior on a monitored computer system and adaptively learns what is normal for individual users, groups, remote hosts, and the overall system behavior. Observed behavior is flagged as a potential intrusion if it deviates significantly from the expected behavior or if it triggers a rule in the expert-system rule base.

Descriptors IDES, intelligent system, intrusion detection, expert system

IDES: A Progress Report

Proceedings of the Sixth Annual Computer Security Applications Conference

Author(s)	Lunt, T.; Tamaru, A.; Gilham, F.; Jagannathan, R.; Neumann, P.; Jalali, C.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: Sixth Annual Computer Security Applications Conference
Publication Date	December 1990
Publisher	Not Known
Subject Treatment	Practical
Document Type	Conference Paper
References	16 Additional References

Abstract This paper describes a real-time intrusion-detection expert system (IDES) that observes user behavior on a monitored computer system and adaptively learns what is normal for individual users, groups, remote hosts, and the overall system behavior. Observed behavior is flagged as a potential intrusion if it deviates significantly from the expected behavior or if it triggers a rule in the expert-system rule base.

Descriptors IDES, intrusion detection, expert system

A Structured Risk Analysis Approach to Resolve the Data Protection and Integrity Issues for Computer-Aided Acquisition Logistics Support (CALS)

Proceedings of the Fifth Annual Computer Security Applications Conference

Author(s)	Gove, R.A.; Friedman, A.R.
Author Affiliation	Booz-Allen & Hamilton, Inc., New York, NY
Citation Source	INSPEC; DIALOG File #13
Document Availability	IEEE Catalog No. 89TH0287-3
Document Source	Proceedings: Fifth Annual Computer Security Applications Conference (pp. 4-5)
Publication Year	1990
Publisher	IEEE Computer Society Press, Los Alamitos, CA
Subject Treatment	Practical
Document Type	Conference Paper
References	No Additional References

Abstract A structured risk analysis approach that is intended to result in cost-effective data protection and integrity service throughout CALS is described. The structured risk analysis approach would: identify CALS assets; determine threats to CALS data; ascertain CALS vulnerabilities; identify potential risks; use the risk and vulnerability assessment as a baseline for protection and integrity identifying the required services; define a generic lattice-ordered set of security labels for unclassified data that will encompass the CALS requirements; develop the specific protocols to implement the architecture; and implement the protocol in a test bed and then conduct security and operational testing.

Descriptors logistics data processing, structured risk analysis approach, computer-aided acquisition logistics support, CALS, lattice-ordered set, security labels

Intrusion and Anomaly Detection in Trusted Systems

Proceedings of the Fifth Annual Computer Security Applications Conference

Author(s)	Winkler, J.R.; Page, W.J.
Author Affiliation	Planning Research Corporation, McLean, VA
Citation Source	Review of Document
Document Availability	IEEE Computer Society
Document Source	Proceedings: Fifth Annual Computer Security Applications Conference
Publication Date	1990
Publisher	IEEE Computer Society Press, Los Alamitos, CA
Subject Treatment	Practical
Document Type	Conference Paper
References	12 Additional References

Abstract Secure systems and networks generate vast amounts of audit information that may reveal unusual situations or patterns of use. While such analysis is usually performed only after other evidence is uncovered, a strong need exists for real-time analysis. The system we describe is a real-time network and host security monitor which allows both interactive and automatic audit trail analysis. Audit records - tokens of actual user behavior, are examined in context of user profiles - measures of expected behavior. This system combines a set of statistical tools for both interactive and automatic analysis of audit data, an expert system that works in conjunction with the statistical tools, and a hierarchical set of audit indicators which are based on an Indications and Warning model. The application of the model allows us to both collect audit events at a fine level of granularity, as well as effectively direct intrusion and anomaly detection by defining levels of concern. A set of discrete tools, capabilities, and components are implemented in a hybrid design utilizing control concepts from operating systems theory and problem-solving concepts from blackboard AI systems.

Descriptors audit trail analysis, behavior profiles, indications and warning model, intrusion detection, anomaly detection

The SRI IDES Statistical Anomaly Detector

Proceedings of the 1991 IEEE Symposium on Security and Privacy

Author(s)	Javitz, Harold S.; Valdes, Alfonso
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: 1991 IEEE Symposium on Security and Privacy
Publication Date	1991
Publisher	IEEE Computer Society
Subject Treatment	Practical
Document Type	Conference Paper
References	1 Additional Reference

Abstract SRI International's real-time intrusion-detection expert system (IDES) system contains a statistical subsystem that observes behavior on a monitored computer system and adaptively learns what is normal for individual users and groups of users. The statistical subsystem also monitors observed behavior and identifies behavior as a potential intrusion (or misuse by authorized users) if it deviates significantly from expected behavior. The multivariate methods used to profile normal behavior and identify deviations from expected behavior are explained in detail. The statistical test for abnormality contains a number of parameters that must be initialized, and the substantive issues relating to setting those parameter values are discussed.

Descriptors intrusion detection, statistical anomaly detection, behavior analysis

Polyinstantiation: An Inevitable Part of a Multilevel World

Proceedings of the Fourth Workshop on the Foundations of Computer Security

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	Unknown
Document Source	Proceedings: Fourth Workshop on the Foundations of Computer Security
Publication Date	June 1991
Publisher	Unknown
Subject Treatment	Practical
Document Type	Conference Paper
References	8 Additional References

Abstract Polyinstantiation is a phenomenon of multilevel data. As such, it exists as a property of information and is not merely the result of any specific technology. Thus, we cannot simply address the recent controversy over the desirability of polyinstantiation in multilevel databases by choosing not to support it in our systems. Rather, we must first recognize that polyinstantiation is an inevitable property of a multilevel world. Once recognizing this, we can then go on to investigate how best to reflect it in our developing technologies. In this position paper, polyinstantiation as a property of the world of multilevel information is discussed.

Descriptors polyinstantiation, muiltilevel security

Abductive and Approximate Reasoning Models for Characterizing Inference Channels

Proceedings of the Fourth Workshop on the Foundations of Computer Security

Author(s)	Garvey, Thomas D.; Lunt, Teresa F.; Stickel, Mark E.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	Unknown
Document Source	Proceedings: Fourth Workshop on the Foundations of Computer Security
Publication Date	June 1991
Publisher	Unknown
Subject Treatment	Practical
Document Type	Conference Paper
References	28 Additional References

Abstract A serious problem in computer database and knowledge base security is detecting and eliminating so-called inference channels. The existence of such channels enables a user with access to information classified at a low level to infer information classified at a high level, and through the transformation of low level data to high level data may provide an unacceptable information flow. Inference channels are particularly difficult to detect automatically because they often involve information that the user has apart from the database and is not accessible in any automated system.

In order to estimate the presence of inference channels, determine the degree of risk which they present, and find ways to eliminate them, we need a formal model to describe them. Here we introduce abductive reasoning, a logical formalism used in artificial intelligence systems for a variety of reasoning tasks. Abduction provides both the basis for a formal model for the inference problem and a computational mechanism for detecting inference channels. Abduction additionally provides a framework for reasoning with approximate and uncertain information which enables us to extend the model for inference channels by taking into account the likelihood that a person might believe some statement of interest.

The methods outlined here have been used for understanding natural language utterances and for diagnosis problems but have not, to the authors' knowledge, been applied to database security. The work described here is very preliminary but, we feel, very promising.

Descriptors inference control, database security, abductive reasoning

Detecting Intruders in Computer Systems

Proceedings of the Sixth Annual Symposium and Technical Displays on Physical and Electronic Security

Author(s)	Lunt, Teresa F.
Author Affiliation	SRI International, Menlo Park, CA
Citation Source	Review of Document
Document Availability	SRI International, Menlo Park, CA
Document Source	Proceedings: Sixth Annual Symposium and Technical Displays on Physical and Electronic Security
Publication Date	Not Known
Publisher	Not Known
Subject Treatment	Practical
Document Type	Conference Paper
References	No Additional References

Abstract This paper describes a real-time intrusion-detection system (IDES) that observes user behavior on a monitored computer system and adaptively learns what is normal for individual users, groups, remote hosts, and the overall system behavior. Observed behavior is flagged as a potential intrusion if it deviates significantly from the expected behavior or if it triggers a rule in the expert-system rule base.

Descriptors IDES, intrusion detection, expert system

APPENDIX B

REQUIREMENTS ANALYSIS REPORT

REQUIREMENTS FOR A CALS DATABASE USAGE ANALYSIS TOOL TO CONTROL AGGREGATION

1.0 BACKGROUND AND OBJECTIVE

The Computer Assisted Acquisition and Logistics Support (CALS) effort will involve large and geographically dispersed databases of proprietary technical information pertaining to weapons systems and parts. These databases will be combined to form the CALS Integrated Weapons Systems Database (IWSDB). While a goal of the CALS effort is to openly exchange information in a standardized format, it is recognized that unlimited access to large amounts of weapons systems data will pose a security risk. These databases will be shared by government and industry personnel and by their very nature, will contain data that, when aggregated, could increase in sensitivity or classification.

The data aggregation problem, as it applies to non-distributed databases, is a current topic of ongoing research in computer database security. Most of the work on aggregation has been concentrated with a few organizations, most notably, SRI. SRI's multilevel relational database system, SeaView, includes work on aggregation and inference, discretionary security, classification constraints, and object-oriented principles. Ford Aerospace Corporation is designing a Database Inference Controller by employing probabilistic knowledge modeling to identify inferences, together with semantic modeling and an expert system component. TRW's work on database security involves again, semantic modeling as well as catalytic data to control what they term "inference aggregation." While this work represents considerable progress toward developing an approach for some aspects of the aggregation problem, there is still much work to be addressed, particularly for distributed databases.

The objective here is to develop the requirements for a tool which will address the aggregation problem. This tool will be referred to as the Data Aggregation Tool (DAT). To keep within the budget for this effort, it was decided to restrict the scope to a single database, leaving the extension to the distributed case for follow-on work.

In the following section, requirements for such a tool are developed by considering the following general questions:

- What is aggregation?
- What is the purpose of the aggregation tool?
- What is the operational philosophy of such a tool?
- How should this aggregation tool fit with or relate to existing intrusion detection technology?
- What constitutes the sphere of influence for the aggregation tool? In other words, how many separate and different sources of information (that could possibly contribute to aggregation) should we worry about?
- What economic issues are important for the development and operation of such a tool?

2.0 DEVELOPING REQUIREMENTS

2.1 WHAT IS AGGREGATION?

Lunt defines aggregation as "... whenever some collection of facts has a classification strictly greater than that of the individual facts forming the aggregate."[1] It is common in the literature to also label the inference problem as aggregation. The definition of the inference problem, again quoting from Lunt [1], is "...whenever some data x can be used to derive partial or complete information about some other data y, where y is classified higher than x." For instance, consider the example where the association of salaries with individual employees is considered secret. Now consider two unclassified lists, one of salaries and one of employees, both indexed by employee number. One could infer the secret salary information from the two unclassified lists.

Here we are interested only in the problem involving a collection of items which, when available together in quantities larger than some threshold N, become classified at a level greater than the individual items, i.e., Lunt's definition of aggregation. This is often referred to as the quantity-based aggregation problem.

The CALS IWSDB will likely encounter this problem in the following context. The various component parts of a particular missile's guidance system may be supplied by several manufacturers. Each manufacturer will have a database of information about each part, including enough information to identify the part as a component of the missile's guidance system. The information about each part, by itself, may be judged to be of only modest value to an enemy, perhaps only requiring a secret classification. If information on all the parts of the guidance system were easily available, however, an enemy might be able to figure out how the system worked and develop a countermeasure for it. Such information would perhaps be classified top secret. This then represents an aggregation problem since assembling all the parts is possible with only a secret clearance while the result could be used to compromise top secret information.

Thus a requirement of the DAT is that it addresses the quantity-based data aggregation problem, i.e., instances where a collection of facts has a higher classification than any one subset of the facts would have individually.

2.2 WHAT IS THE PURPOSE OF THE AGGREGATION TOOL?

There are three levels at which a DAT could operate. At the highest requirement level the purpose of the DAT is to help prevent aggregation. When aggregation cannot be prevented, it should be detected and finally, some action should be invoked to correct the situation in which sensitive data is compromised.

As a tool the focus of DAT should augment and fortify the activities of those responsible for addressing security issues. For databases this would involve the System Security Officer (SSO). The SSO is responsible for ensuring that the design and operation of a database meets security standards. This involves both on and off-line activities. On-line, the SSO must be able to detect security problems as they develop to counter them. Off-line the SSO must ensure that database design meets the necessary requirements for secure operation.

Thus another requirement of the tool is that it be able to be used by the SSO to ensure that the database meets necessary security requirements involving data aggregation and/or to detect data aggregation compromises.

Regarding detection (On-line activities)

There are two approaches for characterizing the behavior of an intruder widely employed by on-line intrusion detection systems: behavioral and scenario. Baur et al [2] indicate that "behavioral characterization works well if specific intrusions cannot be identified, i.e., all user activities are legal and the only concern is that an unauthorized person may be performing them." Behavior modeling takes advantage of the probability that the actions of an unauthorized individual masquerading as an authorized user will have a non-characteristic behavioral pattern. User behavior profiles are kept for comparison with on-going activity. Anyone deviating from usual activity, as defined by their profiles, will be flagged for investigation.

Scenario models rely on a-priori knowledge of the approaches likely to be utilized by an unauthorized individual. A knowledge base is developed for use in comparing current user activities to predetermined patterns or scenarios of unacceptable or intrusive behavior.

This leads to the further requirement that the DAT contain a knowledge base or interface with an external knowledge-based system which may contain both static and dynamic knowledge bases capable of learning new user behavior patterns as they change naturally over time.

Regarding design (Off-line activities)

The design of a DAT should proceed in conjunction with the design of the IWSDB. By doing so, we will be able to build-in security mechanisms that can help to prevent aggregation. A good database design will facilitate the implementation of a security policy. A good tool design should take advantage of the architecture of the database and the constructs that comprise it.

As an example of the synergism possible between a DBMS and a DAT, consider a DBMS based on the object-oriented model. The object model can support a hierarchy of classes of objects. Relationships can be defined between any two levels of the hierarchy. One such example is the "is-part-of" relationship. Each such object may be comprised of facets which relate it to other objects of which it is composed. Thus a facet of an object A may identify another object B which "is-part-of" object A. Object B might, in turn, have facets which relate it to other objects of which it is composed, and so on.

Lunt [6] mentions that objects related through the "is-a" relationship can be protected from aggregation by enforcing a rule that the classification of a facet must dominate that of its object.

With this rule we can allow objects in a subclass to have a lower classification than those in their parent class. This "is-a" relationship that connects the two classes would have a classification at least as high as the parent object and would be part of a mechanism to protect the parent object from uncleared users. Indeed, individuals accessing the subclass who lack the necessary clearance to access the parent class would not even know of the relationship to the parent class.

In the case of the IWSDB, it would seem more appropriate to adapt this rule; i.e., that a facet's classification must dominate that of its object to a rule containing the "is-part-of" relationship. In this way many parts of a classified system can be made available to users (because the parts themselves are unclassified), and only those cleared to the appropriate level will be aware of the fact that the parts belong to the classified system.

Thus a DAT which facilitates the establishment of object-oriented database models, including the enforcement of constraints such as mentioned here for the facets, will help prevent aggregation.

This leads to the requirement that a DAT be designed in conjunction with the design of the IWSDB. In addition, it seems as though a DBMS based on the object-oriented model is well suited to handle the aggregation problem.

2.3 WHAT IS THE OPERATIONAL PHILOSOPHY OF SUCH A TOOL?

Autonomous vs. Man-in-the Loop Considerations

No IDS system to date is capable of operating in a stand-alone mode. All systems currently being designed are intended to work in conjunction with a human SSO. Given the complexity of recognizing all potential security problems along with the importance of ensuring that no compromises exist, it is very unlikely that any computer-based system would be allowed to operate in a completely autonomous mode. *Thus another requirement will be that the DAT operate as a tool to be used by the SSO.* The tool may make recommendations, but no final decisions should be expected of the tool.

An acceptable operational philosophy would probably have DAT draw on audit data obtained from the database and consult with the expert system component to classify an aggregate. The classification along with the reasoning leading up to it would be made available for the SSO to

examine. Further data can then be obtained if the SSO considers it necessary, or action can be taken to deny further access to the intruder.

An SSO may also be required to sanitize information before it is released to a user. As Lunt et al [3] suggest, this may involve restricting the number of facts per period that are released to a user. *This leads directly to the requirement that the DAT must have a system high interface to an SSO.*

If DAT flags a potentially sensitive aggregate, the SSO may want more detailed audit data on that process to help determine whether there is a security risk. *Thus a variable audit capability, similar to what Hubbard discusses in [4], should be a requirement of DAT.*

2.4 HOW SHOULD THIS AGGREGATION TOOL FIT WITH OR RELATE TO EXISTING INTRUSION DETECTION TECHNOLOGY?

The most obvious means of intrusion detection is to detect suspicious behavior, i.e., behavior that indicates a user is accessing system resources in an atypical fashion. While computer systems produce prodigious amounts of audit data on system activity, the data is not formatted, making it virtually impossible for a system security officer (SSO) to detect any kind of misuse. This gave rise to the development of Intrusion Detection Systems (IDS) in the form of audit trail analysis tools and intrusion detection expert systems. IDS systems are used to aid the SSO in the identification of unauthorized use of a computer system. In fact, audit trail or usage analysis is the foundation for all IDS systems currently in development.

It is not the purpose of the DAT to detect intruders, but rather to detect aggregation. However, much of the information produced by the IDS may be useful for that component of DAT addressing aggregation detection. *To this end a further requirement is that the DAT should utilize the audit data provided by the IDS.*

2.5 WHAT CONSTITUTES THE SPHERE OF INFLUENCE FOR THE AGGREGATION TOOL?

In other words, how many separate and different sources of information (that could possibly contribute to aggregation) should be considered?

The aggregation problem is still more complex than has been alluded to. Seemingly innocuous aggregates could be combined with outside knowledge to comprise a classified aggregate. "Outside knowledge" refers to data that is not in the Integrated Weapons System Database (IWSDB) but may be known to the user by some other means. Morgenstern refers to this collection of outside knowledge as a "Sphere of Influence." "A Sphere of Influence (SOI) models the process by which a user's knowledge of an application can give rise to inferences about additional information." [5]

Although controlling aggregation based on SOI knowledge is beyond the scope of this effort, DAT "must have available to it relevant knowledge of the application" [5] to identify those data items that are not sensitive to aggregation. "Such knowledge must make explicit the constraints that cause different data values to be interdependent." [5] This knowledge may exist in the heads of SSOs or sophisticated IWSDB users. It is important that the help of such individuals be elicited as this information will help to limit the complexity of the knowledge base.

DAT should facilitate an exchange of information between the sophisticated IWSDB user and the SSO to help the SSO classify aggregates.

2.6 WHAT ECONOMIC ISSUES ARE IMPORTANT FOR THE DEVELOPMENT AND OPERATION OF SUCH A TOOL?

To remain economically feasible, DAT must be able to accept as input the standard audit trail information generated by the host operating system and processed by an IDS system and must be able to interface to a commercial DBMS. DAT must balance cost with goals. The amount of processing required must not exceed the benefits based on accepted confidence limits. *To this end, DAT should operate in polynomial time.*

3.0 CONCLUSIONS

This paper begins to address the requirements for a Data Aggregation Tool (DAT). Detecting aggregation is, in many respects, analogous to detecting intruders. Much research has been done on Intrusion Detection Systems (IDS) and the design of a DAT should build on this work.

The design of a DAT should take into consideration the *prevention* of aggregation as well as *detection* and recovery from compromise. Towards prevention, a DAT should be capable of protecting sensitive relationships between data items. Towards detection, the tool should operate in polynomial time and support a variable audit capability. With regards to recovery, a DAT should include a learning element that will derive new security rules to prevent a compromising scenario from recurring. To this end, the DAT should be designed in conjunction with the IWSDB to ensure a synergistic relationship. An object-oriented DBMS holds much promise for handling the aggregation problem.

Interpreting security policy, while taking into consideration world events and temporal constraints, is an extremely complex task; one that cannot be entrusted entirely to a computer-based system. A DAT should be designed to work with a human SSO to best ensure that no classified data is compromised.

Relevant security policies must be considered in order to best protect the data that needs protecting without encumbering non-sensitive data.

REFERENCES

[1] Lunt, Teresa F.,"Aggregation and Inference: Facts and Fallacies," Proceedings of the 1989 IEEE Symposium on Research in Security and Privacy, Oakland, CA, May 1989.

[2] Baur, D.S.; Eichelman, F.R. II; Herrera, R.M.; Irgon, A.E.; "Intrusion Detection: An Application of Expert Systems to Computer Security," 1989

[3] Lunt, Teresa F.; Neumann, Peter G.; Denning, Dorothy; "Secure Distributed Data Views, Vol. 1: Security Policy and Policy Interpretation for a Class A1 Multilevel Secure Relational Database System," 1988

[4] Hubbard, Brian et al, "Computer System Intrusion Detection," 1990

[5] Morgenstern, Matthew, "Intelligent Database Systems," 1990

[6] Lunt, Teresa F., "Multilevel Security for Object-Oriented Database Systems," Proceedings of the Third IFIP WG 11.3 Workshop on Database Security, Monterey, CA, September 1989